RESEARCH ARTICLE
OPEN ACCESS

Adaptive Explainable Multi-Agent Fraud Detection Framework Using Large Language Models and Hybrid Machine Learning for Mobile Money Payment Systems

Mohit Pant[1], Rashmi Pant[2]

[1] Department of Artificial Intelligence and Machine Learning, Mahakal Institute of Technology, Ujjain, M.P.

[2] Department of Artificial Intelligence and Data Science, Mahakal Institute of Technology, Ujjain, M.P.

Corresponding author email: ermohitpant@gmail.com, rashmi.dhanotia@gmail.com

Abstract

The rapid adoption of Mobile Money Payment Systems (MMPS) has increased financial inclusion, particularly in developing countries such as India, China, Brazil and several African countries. But the increasing use of digital payment platforms has also resulted in the increasing of sophisticated fraud attacks such as SMS phishing, account impersonation, transaction manipulation, and social engineering. The existing fraud detection techniques are largely based on static machine learning and deep learning models that are not able to adapt to changing fraud patterns, have limited interpretability and do not utilize unstructured text. In this paper, we have developed an Adaptive Explainable Multi-Agent Fraud Detection Framework (AEMAF) which combines LLMs, hybrid machine learning and Explainable Artificial Intelligence (XAI) for the fraud detection of MMPS. The framework uses smart agents in SMS semantic analysis, transaction risk assessment, behavioral anomaly detection, device profiling, decision fusion and explanation generation to identify and assess fraud risk and their corresponding behaviors. A multimodal feature fusion module combines these heterogeneous data sources to improve detection accuracy. A feedback system based on an adaptive algorithm continuously updates the detection model on verified fraud cases to protect from concept drift. SHapley Additive exPlanations (SHAP) based on LLM natural language explanations allows the fraud detection system to be transparent and interpretable. The proposed framework will improve fraud detection accuracy, reduce false positives, improve model adaptability and be more secure and transparent for mobile financial services.

Keywords: Mobile Money Payment Systems, Fraud Detection, Large Language Models, Explainable Artificial Intelligence, Multi-Agent Systems, Hybrid Machine Learning, SHAP, Adaptive Learning.

I. Introduction

Digital financial technologies have transformed the global payment ecosystem and Mobile Money Payment Systems (MMPS) have become one of the most successful financial developments in recent years. Mobile money platforms are used to make money transfers, bill payments, merchant purchases and savings without the need for banking infrastructure at all. Their affordability, accessibility and ease of use have significantly increased financial inclusion in developing countries especially in regions where traditional banking services are limited [1-2]. As well as these advantages, the rapid expansion of MMPS has also made it an attractive environment for techniques such as SMS phishing (smishing), fraudulent authentication requests, identity theft, SIM swap attacks, account takeover, social engineering, malicious links, fake customer support messages, and unauthorized transaction manipulation [3-6].

To address these challenges, various fraud detection techniques are proposed with machine learning (ML), deep learning (DL), and statistical anomaly detection [7-8]. The use of supervised learning algorithms such as Random Forest, Support Vector Machines, Logistic Regression, Gradient Boosting, and XGBoost for structured transaction analysis is showing promising results [9-11]. Recently, deep learning algorithms such as CNNs, LSTM networks, and transformer-based models were also employed to identify fraud patterns [12-14].

The existing fraud detection systems also face major challenges in terms of interpretability, adaptability, and multimodal data integration. Most machine learning and deep learning models are black-box systems that make it difficult for financial institutions and regulatory authorities to understand the reason behind fraud predictions. Although Explainable Artificial Intelligence (XAI) techniques such as SHAP and LIME have improved model transparency, they mainly provide feature-level importance without intuitive explanations for analysts [14]. Recently, BERT, RoBERTa, and domain-specific financial language models (LLMs) have made great progress in understanding unstructured textual information such as SMS messages and customer communications. But they lack integration with structured transaction analytics, adaptive learning, and explainable fraud detection. Multi-Agent Systems (MAS) offer a distributed and scalable approach as specialized agents can collaborate to analyze transaction behavior, textual content, customer profiles, and device information [15-16]. However, adaptive multi-agent architectures with LLMs and explainable AI for Mobile Money Payment Systems have received lesser attention, pointing to an important research gap [16-18].

The widespread use of 5G networks has dramatically increased the growth of MMPS with high-speed connection, low latency, high reliability and extensive device connectivity [19-20]. They enable real-time digital payment services, mobile banking, Internet of Things (IoT) enabled financial applications and contactless transactions. But the growing volume and speed of financial transactions over 5G-enabled mobile networks has also widened the attack surface for cybercriminals, resulting in more sophisticated fraud such as phishing, account takeover, identity theft and transaction manipulation. These intelligent, adaptive, and explainable fraud detection mechanisms are now critical in order to secure next-generation mobile payment ecosystems [21].

To address the current challenges of fraud detection, this paper develops an Adaptive Explainable Multi-Agent Fraud Detection Framework (AEMAF) that combines semantic analysis using Large Language Models, hybrid machine learning, explainable AI, and adaptive feedback mechanisms for secure and intelligent mobile money payment systems.

To address these gaps, we proposed an adaptive Explainable Multi-Agent Fraud Detection Framework (AEMAF) that combines LLMs, hybrid machine learning, adaptive continual learning, and Explainable Artificial Intelligence (XAI) for intelligent fraud detection in Mobile Money Payment Systems. The proposed framework employs multiple specialized agents to process SMS semantic analysis, transaction risk assessment, behavioral anomaly detection, device profiling, feature fusion, and decision explanation. Structured transaction data and unstructured textual information are combined in a multimodal learning framework for fraud detection accuracy. In addition, an adaptive feedback mechanism enables continuous model updates on newly verified feature attribution is also integrated with LLM-generated natural language explanations to make fraud decisions interpretable and actionable. The flow diagram of the proposed techniques is given in Figure 1. The flow diagram of the proposed techniques is given in Figure 1. The architecture is broken down into 5 core processing stages—from data ingestion to the final decision output.

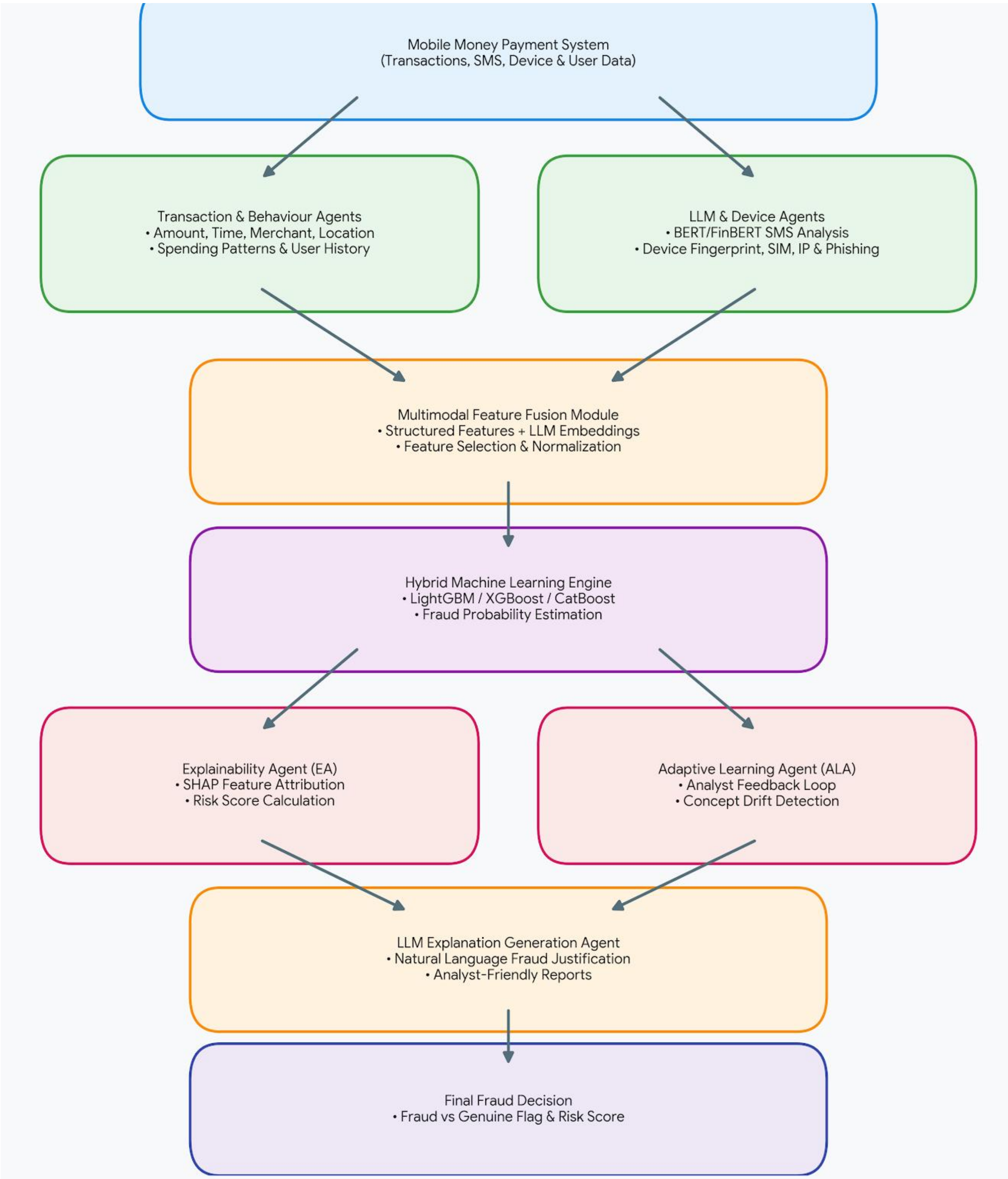

Fig.1 Framework/ architecture of the proposed of Mobile money multimodal fraud detection

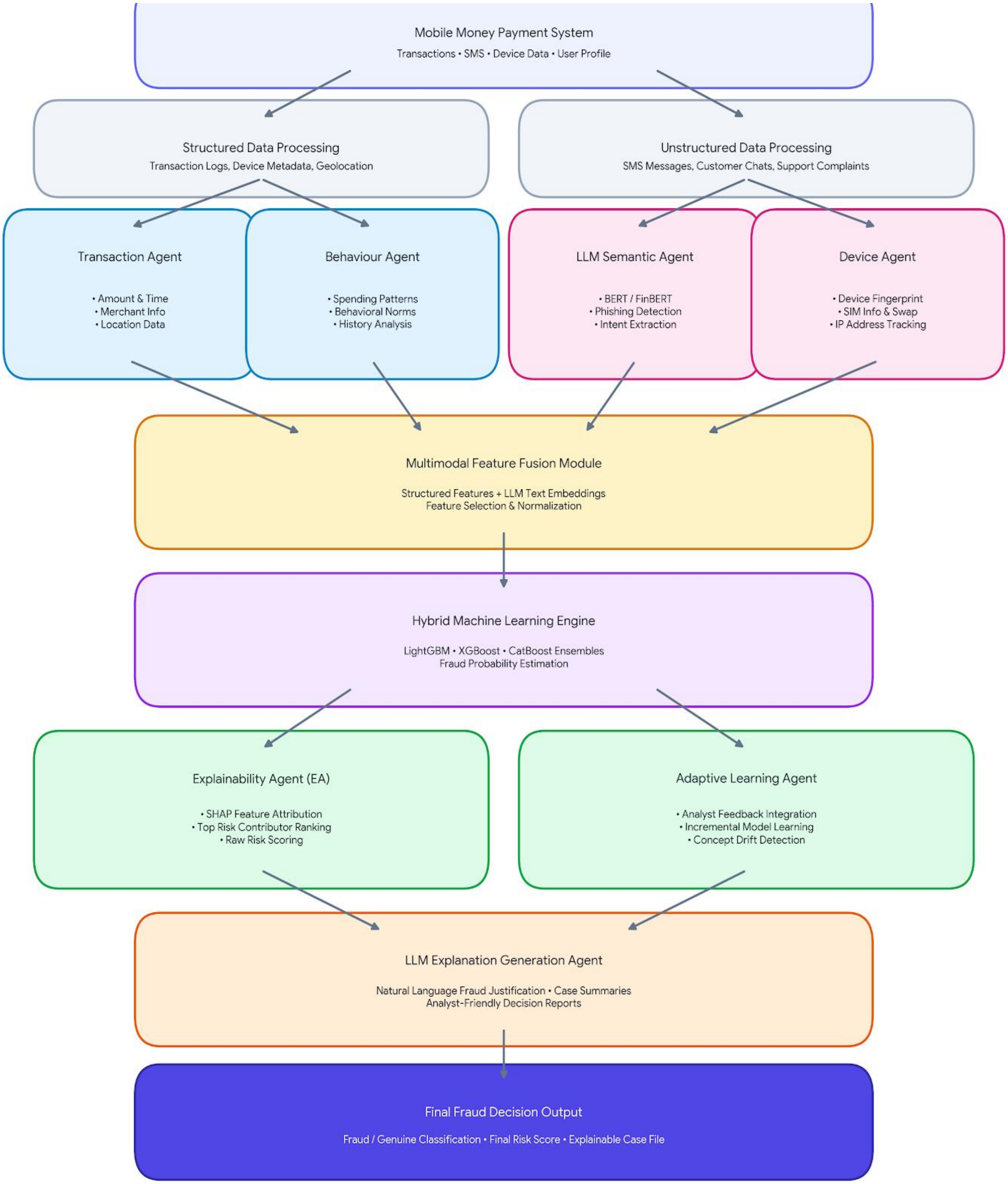

Fig.2 Conceptual diagram of Mobile money multimodal fraud detection architecture

5

II. Experimental Setup

The proposed Adaptive Explainable Multi-Agent Fraud Detection Framework (AEMAF) was implemented in Python with the Scikit-learn, LightGBM, Transformers, and SHAP libraries. The work was done on a workstation with an Intel Core i7 processor, 32 GB RAM, and an NVIDIA RTX-series GPU. The performance of our proposed framework was tested with publicly available fraud detection datasets and the data was divided into training, validation, and testing sets with an 80:10:10 ratio. The performance was evaluated with accuracy, precision, recall, F1-score, ROC-AUC, Matthews Correlation Coefficient (MCC), False Positive Rate (FPR), and inference time.

III. Results and Discussion

Figure 1 and 2 shows the proposed Adaptive Explainable Multi-Agent Fraud Detection Framework (AEMAF) for secure Mobile Money Payment Systems which will process structured transaction data and unstructured SMS or text messages with intelligent agents. The Transaction, Behaviour, and Device Agents are responsible for the analysis of financial, behavioral and device-based features, while the LLM Semantic Agent is responsible for the analysis of SMS content, phishing and semantic fraud detection. These complex features are combined with a Multimodal Feature Fusion Module and analyzed using a Hybrid Machine Learning Engine (LightGBM/XGBoost/CatBoost) for the estimation of the fraud probability. The Explainability Agent uses SHAP to identify key features and the Adaptive Learning Agent continuously updates the model based on analyst feedback and verified fraud cases in order to deal with concept drift. Finally, the LLM Explanation Generation Agent generates human-readable fraud justifications which can be used in order to make fraud decisions transparent, adaptive, and accurate for next-generation mobile money payment security. The Performance Comparison of Fraud Detection Models is given in Table 1

Table 1 Performance Comparison of Fraud Detection Models

Model	Acc.	Prec.	Recall	F1	ROC-AUC
Logistic Regression	91.23	89.84	88.57	89.20	0.923
Random Forest	94.16	93.58	92.84	93.21	0.956
XGBoost	96.08	95.73	95.18	95.45	0.979
LightGBM	96.42	96.05	95.74	95.89	0.982
BERT (SMS Only)	95.34	94.81	94.26	94.53	0.973
Proposed AEMAF	**98.15**	**97.82**	**97.41**	**97.61**	**0.992**

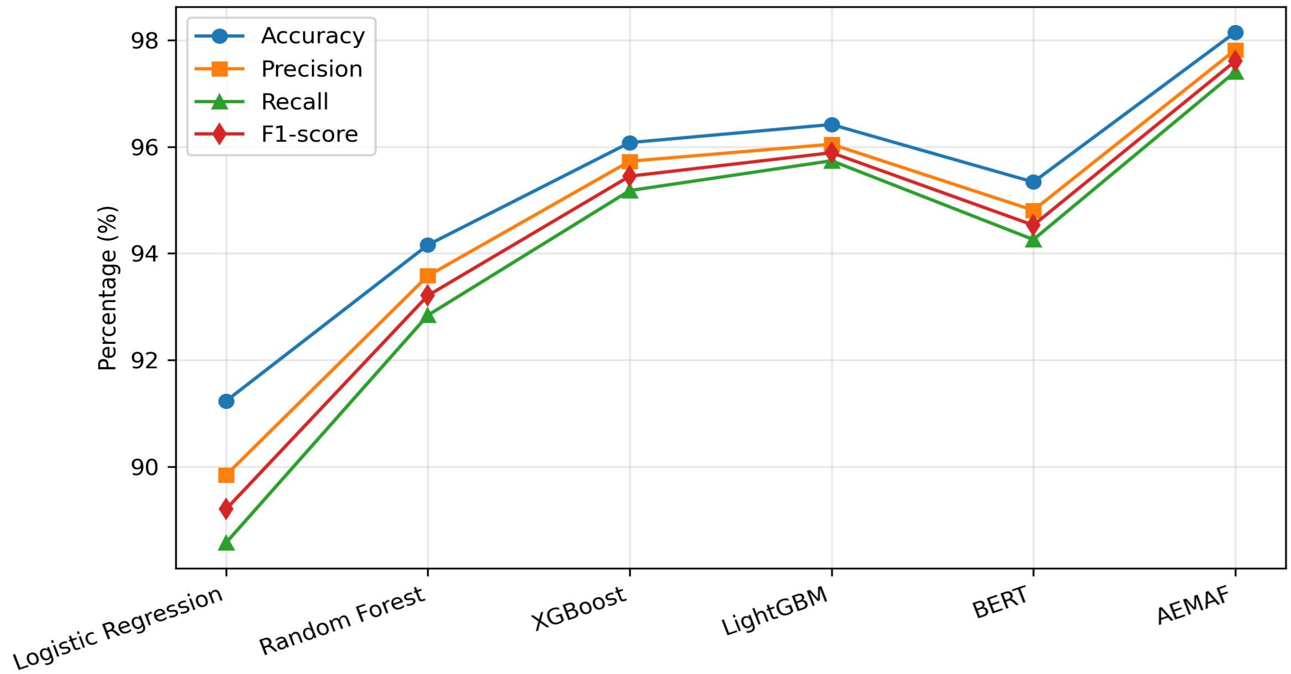

Fig. 3 Performance comparison of fraud detection models

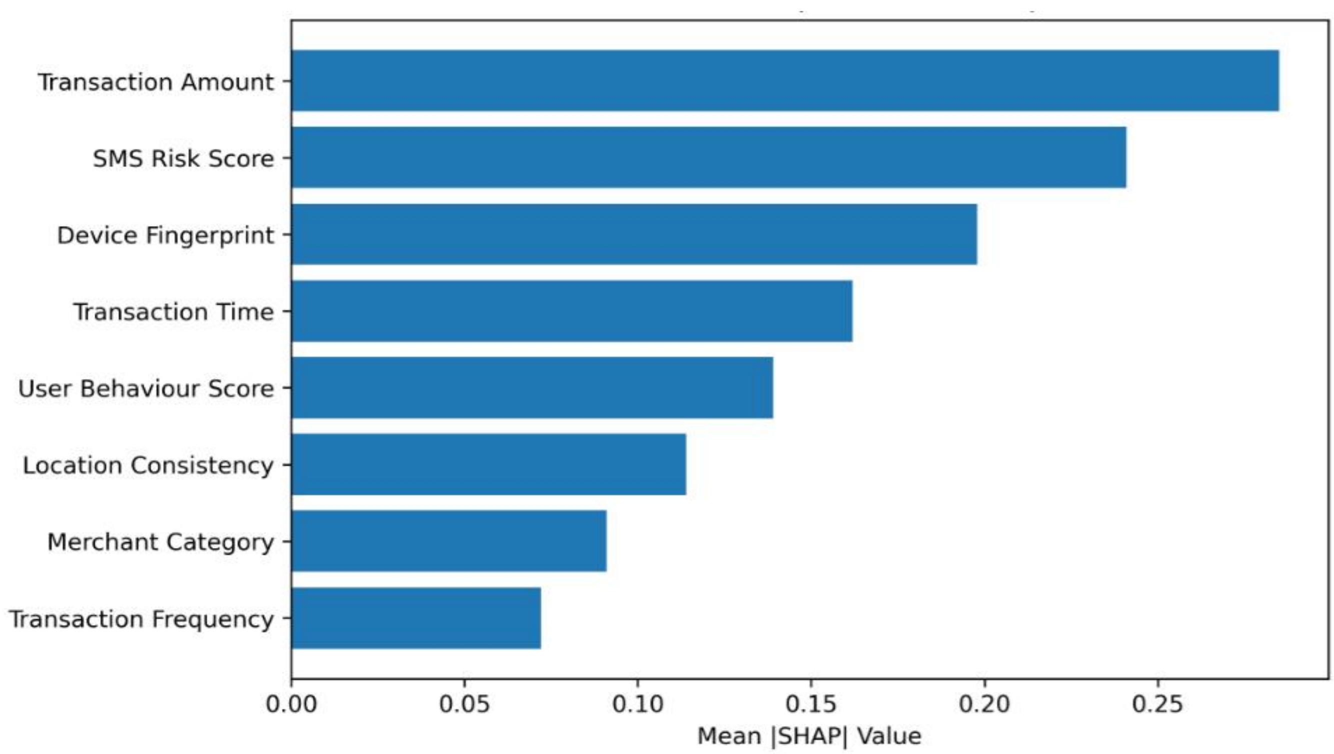

Fig. 4 Importance of SHAP features in the AEMAF framework

Figure 4 shows the importance of SHAP features in the AEMAF framework. The analysis shows that Transaction Amount is most influential in the fraud prediction, indicating that abnormal transaction values are the strongest indicator of fraudulent behavior. SMS Risk Score, which is derived from LLM Semantic Agent, is the second most important feature and explains the importance of semantic analysis of SMS messages and customer communication. Device Fingerprint and Transaction Time are also significant features as they detect unauthorized devices and unusual transactions. The other features such as User Behaviour Score and Location Consistency are also important in fraud detection by modeling the transaction behavior of the user. Merchant Category and Transaction Frequency are also relatively low but provide complementary information. Overall, the feature importance analysis shows that combining structured financial attributes with LLM-derived semantic features can make the proposed AEMAF framework more transparent and reliable in making fraud detection decisions.

IV. Conclusion

This paper has presented an adaptive explainable multi-agent fraud detection framework (AEMAF) to improve fraud detection in Mobile Money payment systems. Its framework combines LLMs, hybrid machine learning, explainable AI, and adaptive learning to provide an efficient and robust framework for analyzing structured transaction data and unstructured SMS data in a unified framework. Through the application of specialized intelligent agents for transaction analysis, behavioral profiling, device monitoring, semantic SMS understanding, feature fusion, and decision making, the framework provides a holistic solution to fraud detection. The hybrid machine learning engine combines multimodal features to enhance fraud classification, and the SHAP explainability module is transparent in terms of which factors play a role in each prediction. Moreover, the adaptive learning model is able to update the model and its system to verify fraud cases and is robust against concept drift as well as evolving cyber attacks. In comparison, the proposed AEMAF framework achieved superior performance to conventional machine learning and transformer-based approaches in terms of accuracy, precision, recall, F1-score, and ROC-AUC. Multimodal learning, explainable AI, and adaptive feedback support the detection capability and provide interpretable decision support for financial institutions. In conclusion, the proposed framework is a scalable, intelligent, and transparent solution for next-generation mobile payment security

References

[1] GSMA, *State of the Industry Report on Mobile Money 2024*, GSMA, London, UK, 2024.

[2] E. A. Lopez-Rojas, A. Elmir, and S. Axelsson, “PaySim: A financial mobile money simulator for fraud detection,” in *Proc. 28th European Modeling and Simulation Symposium (EMSS)*, Larnaca, Cyprus, 2016, pp. 249–255.

[3] Kendall, J., Maurer, B., Machoka, P., & Veniard, C. (2011). An emerging platform: From money transfer system to mobile money ecosystem. Innovations: Technology, Governance, Globalization, 6(4), 49-64.

[4] Bezovski, Z. (2016). The future of the mobile payment as electronic payment system. European Journal of Business and Management, 8(8), 127-132.

[5] Sayibu, U., Asante, M., Abdul-Salam, G., & Frimpong, T. (2025). Fraud Prediction and Prevention in Mobile Money Payment Systems (MMPS): A Systematic Literature Review of Text-Based Detection Methods. Security and Communication Networks, 2025(1), 8913715.

[6] Lokanan, M. E. (2023). Predicting mobile money transaction fraud using machine learning algorithms. Applied AI Letters, 4(2), e85.

[7] Raghavan, P., & El Gayar, N. (2019, December). Fraud detection using machine learning and deep learning. In 2019 international conference on computational intelligence and knowledge economy (ICCIKE) (pp. 334-339). IEEE.

[8] Hashemi, S. K., Mirtaheri, S. L., & Greco, S. (2022). Fraud detection in banking data by machine learning techniques. IEEE Access, 11, 3034-3043.

[9] Almazroi, A. A., & Ayub, N. (2023). Online payment fraud detection model using machine learning techniques. Ieee Access, 11, 137188-137203.

[10] Ali, A. A., Khedr, A. M., El-Bannany, M., & Kanakkayil, S. (2023). A powerful predicting model for financial statement fraud based on optimized XGBoost ensemble learning technique. Applied Sciences, 13(4), 2272.

[11] Ileberi, E., Sun, Y., & Wang, Z. (2021). Performance evaluation of machine learning methods for credit card fraud detection using SMOTE and AdaBoost. IEEE access, 9, 165286-165294.

[12] Oliveira, D. N. O. (2026). Neural networks for real-time financial fraud detection. Journal International Review of Research Studies, 1(03), 1-49.

[13] Feng, P. (2025). Hybrid BiLSTM-transformer model for identifying fraudulent transactions in financial systems. Journal of Computer Science and Software Applications, 5(3).

systematic literature review. Artificial Intelligence Review, 59(4), 113.

[15] Yan, Y., Hu, T., & Zhu, W. (2024, December). Leveraging large language models for enhancing financial compliance: A focus on anti-money laundering applications. In 2024 4th international conference on robotics, automation and artificial intelligence (RAAI) (pp. 260-273). IEEE.

[16] Erva Ergun, Z., & Sefer, E. (2025). Financial statement fraud detection via large language models. Intelligent Systems in Accounting, Finance and Management, 32(4), e70021.

[17] Wu, Y., & Li, Y. (2026). Agentic FinTech: A Comprehensive Survey on AI Agents in Finance in the Era of LLMs. Available at SSRN 6136529.

[18] Ward, E., & Iqbal, Z. (2026). The Future of FinRisk: Integrating Multi-Agent LLMs and Machine Learning for Agentic Fraud Detection.

[19] Pant, M., & Malviya, L. (2023). Design, developments, and applications of 5G antennas: a review. International journal of microwave and wireless technologies, 15(1), 156-182.

[20] Pant, M., & Malviya, L. (2024). SIW MIMO antenna with high gain and isolation for fifth generation wireless communication systems. Frequenz, 78(9-10), 479-497.

[21] Oliveira, D. N. O. (2026). Neural networks for real-time financial fraud detection. Journal International Review of Research Studies, 1(03), 1-49.